Perspectives

Anecdotal, Historical and Critical Commentaries on Genetics

Edited by James F. Crow and William F. Dove

MAPPING FUNCTIONS

IT is said that R. A. FISHER, on first meeting the MORGAN group of Drosophila geneticists, asked, "Why don't you people ever do a proper mapping experiment?" This, to the group that invented linkage maps, seems like asking Shakespeare or Shaw why he never learned to use the English language. I don't know whether the anecdote is true, but if it is, one can easily guess that FISHER had three things in mind.[1] First, the MORGAN group paid little attention to estimating statistical errors. Second, they never did the kind of balanced experiment that FISHER advocated, *i.e.*, using two matings in which the heterozygous parents are in opposite linkage phases, so that viability differences cancel (although MULLER thought of it). Third, they ordinarily made no distinction between the percent of recombination and map distance.

The Drosophila group did experiments involving large numbers under fairly standardized environmental conditions, which reduced statistical errors. But the mutant phenotypes were almost always less viable than wild type. When viability differences were large, the experimenters made arbitrary corrections, or devised special tricks. For example, they could keep deleterious recessives heterozygous and identify genotypes by progeny tests, although this was extremely laborious. A serviceable alternative, if interactions were small, was BRIDGES' procedure of alternating mutant and wild-type alleles along the chromosome. Finally, the strategy was to build up the map by combining distances between adjacent genes, and as the map became more dense, undetected double crossovers were decreasingly important. So FISHER's refinements, although they might have improved efficiency, made little difference in the long run.

FRANK STAHL (1989) referred to these two approaches as the American and the British. The American style was to use brute force, collecting so many mutant genes that the intervals between adjacent pairs were small. The cunning British used a mapping function to correct for undetected multiple crossovers. Thus, fewer markers were needed and the lesser effort "leaves time for tea."

Balanced experiments to correct for interactions among viability factors are particularly neat. Suppose two matings are made in equal numbers, (1) $AB/ab \times ab/ab$ and (2) $Ab/aB \times ab/ab$, and let R_1, N_1, R_2 and N_2 be the expected proportion of recombinants and nonrecombinants in the two experiments. Suppose further that the viability of $Ab + aB$ phenotypes is v relative that that of $AB + ab$. Then the expected ratio of recombinants, r, to nonrecombinants, $1 - r$, in the combined experiments is $\sqrt{R_1 v R_2 / N_1 N_2 v}$ and the viability factor, v, cancels out regardless of interactions. Equating the corresponding function of the observed proportions, $\sqrt{R_1 R_2 / N_1 N_2}$, to $r/(1 - r)$ gives a consistent estimate of r. A simple formula for the variance of r is available (FISHER 1949a, p. 221*ff*). The two matings also permit direct estimates of the viabilities. Extension to three factors is straightforward, but four balanced matings are required. Such an experiment was carried out in mice by MARGARET WALLACE (see FISHER 1949b). In a way the result was an anticlimax, because there were no detectable viability differences. Such experiments have not been the practice in Drosophila. One reason is that most Drosophila mapping experiments involve multiple markers, and the number of matings required for a balanced experiment doubles with each added marker.

The MORGAN school and its successors throughout the Drosophila world have not made use of mapping functions. The group was, of course, keenly aware that there were multiple exchanges and that the proportion of recombinants was not linear with map distance when the distances were large; and MULLER had defined coincidence and interference. In the early days, before the map became dense, corrections for undetected multiple crossovers would have been useful. HALDANE's (1919) pioneering paper was ignored, however.

Following HALDANE's procedure, let r be the pro-

[1] I would enjoy hearing from anyone who can offer any information as to the truth of this story.

portion of recombinants and m the map distance in morgans. His formula is then

$$\frac{dr}{dm} = 1 - 2cr \qquad (1)$$

in which c is the coincidence, (actual double crossovers)/(number expected with no interference). Incidentally, this paper also added to the vocabulary of genetics: here HALDANE proposed "that the unit of distance in a chromosome . . . be termed a 'morgan,' on the analogy of the ohm, volt, etc. Morgan's unit of distance is therefore a centimorgan."

HALDANE noted that when $c = 0$, as is true for closely linked genes, then

$$r = m \qquad (2)$$

and recombination percent and map distance are the same. He also noted that as the distance increases, c approaches 1. When $c = 1$, we get the well known HALDANE mapping function

$$r = \tfrac{1}{2}(1 - e^{-2m}). \qquad (3)$$

It is not fair to HALDANE to say that he thought of this as a realistic mapping function, for in this paper he showed that the already quite extensive Drosophila data fell between the curves given by (2) and (3). A possible reason for the nonuse of mapping functions may be that the function that HALDANE found to best fit the data was a very complicated one, with no theoretical underpinning.

Not for a quarter century (KOSAMBI 1944) did a formula appear that caught on. A natural extension of constant c in Equation 2 is to let it be a function of r. KOSAMBI took the simplest and very reasonable next step, making $c \propto r$. Letting $c = 2r$, which gives the correct values at $r = 0$ and 1, leads immediately to the widely used KOSAMBI mapping function

$$r = \tfrac{1}{2}\tanh 2m. \qquad (4)$$

This formula fits most data fairly well, or at least well enough for most purposes. I think one reason for its popularity is that tables of hyperbolic functions have long been readily available, and in precomputer days they saved a great deal of tedious calculation. Furthermore, the addition rule for adding recombination values follows from the ordinary rules of hyperbolic functions

$$r_{12} = \frac{r_1 + r_2}{1 + 4r_1r_2}. \qquad (5)$$

Devising different mapping functions has provided diversion, if not gainful employment, for a number of mathematically inclined geneticists. Most of the earlier work in this area has been reviewed by BAILEY (1961). Equation 1 has natural extensions in several directions, of which I shall mention three. FELSENSTEIN (1979) assumed that c increases linearly with r, but takes a value K when $r = 0$. This has the advantage of permitting "map expansion" ($K < 0$) and "negative interference" ($K > 1$). Like the KOSAMBI function it has a simple addition rule, analogous to (5). Its flexibility, which may be desirable for some situations, can be undesirable for routine mapping in higher eukaryotes if one wants $c = 0$ for short distances; in this case the FELSENSTEIN and KOSAMBI formulas are the same. For a lucid and entertaining discussion of mapping functions in prokaryotes and references to earlier work, see STAHL (1989).

A second, and natural, extension is to let c be proportional to a higher power of $2r$. CARTER and FALCONER (1951) assumed that $c = (2r)^3$ and preferred this for mouse data. A more elaborate formula by RAO *et al.* (1977) includes the KOSAMBI and CARTER-FALCONER formulas as special cases. It has one adjustable parameter and, when this is chosen optimally, the function gives a very good fit to data on human chiasma frequencies.

PASCOE and MORTON (1987) found that the formula of RAO *et al.* gave a better fit to the Drosophila data than any of the others that I have mentioned. They noted, however, that a simpler formula assuming $c = (2r)^2$ gave essentially the same results. The integrated form is given as Equation 3 in their paper. Given that the data lie between the equation with $c = 2r$ and $c = (2r)^3$, then, as PASCOE and MORTON implied, it requires no great intellectual leap to think of $(2r)^2$. In fact I had used this in an elementary textbook (see curve C on p. 68 of CROW 1983).

The analysis of PASCOE and MORTON (1987) is reassuring in that the best fitting formulas are the same for data based on human chiasma frequencies as for recombination data in Drosophila. Furthermore, the high interference over the range 0–15 cm in the mouse (KING *et al.* 1989) is consistent with $c = (2r)^2$, although the data are insufficient to distinguish between this and the CARTER-FALCONER function.

In contrast to these more or less empirical functions, there are those that depend on specific, more mechanistic assumptions. One such is the model of FISHER (1948) and OWEN (1950), in which exchanges start at some point, say the centromere, with an assumed probability distribution, and subsequent exchanges depend on this and on an interference function. One feature of this model is that it permits greater than 50% recombination for certain map distances. Some mouse data support this, but the question warrants further investigation.

All of these functions overpredict the number of triple and higher crossovers, suggesting higher-order interference not accounted for in theories developed from three-locus models. This is strikingly evident in

7-point and 9-point crosses in Drosophila (PASCOE and MORTON 1987). Furthermore, it is to be expected that there are telomere and centromere effects as well as local differences in coincidence, so that no global mapping function can be correct everywhere. An approach based on chiasma distributions has been advocated by GOLDGAR, FAIN and KIMBERLING (1989). Chiasmata serve two functions. One is extending the evolutionary advantages of recombination to genes on the same chromosome. The other is mechanical, regularizing meiosis by reducing nondisjunction. These functions may call for different numbers and distributions of chiasmata, and the existing values may be some sort of compromise between conflicting optima. In any event, the chiasma distribution is far from random, and taking this into account may well provide a better approach to mapping.

The increasing practicality of human multipoint mapping, foreshadowed by BOTSTEIN *et al.* (1980) and involving increasingly effective computer routines (LANDER and GREEN 1987; LATHROP and LALOUEL 1988), raises anew the question of mapping functions. To write likelihood equations, one must make some assumption about interference. It is not clear (to me, at least) how much practical difference the specific interference assumption makes (see for example LATHROP *et al.* 1985 and PASCOE and MORTON 1987). In any case it is comforting that the simple formula based on $c = (2r)^2$ works well in those circumstances in which the data are sufficient to test it, and it can be built into computer routines. To quote GOLDGAR, FAIN and KIMBERLING (1989), "Given the enormous resources now being devoted to mapping and sequencing the entire human genome, methods which produce even a modest increase in efficiency are of value."

During the long period when human linkage studies were making very little progress for want of sufficient markers, the mouse map was progressing steadily. The pioneering role of J. B. S. HALDANE and L. C. DUNN in getting this started, and the progress and improved methodology in the ensuing years, was reviewed in this column two months ago (LYON 1990). Detailed comparative mapping of murine and human genes, now at hand, offers exciting prospects.

Linkage maps are a step toward physical maps. But they are more. However detailed the physical map, down to knowing the nucleotide sequence, we require linkage information in many species for the study of transmission genetics and for analyzing phenotypes whose molecular basis is not known. Human genetics, specifically, calls for ever better means of predicting the gametic output of persons of specified genotype. For this, chromosome maps and interference functions will continue to be needed.

JAMES F. CROW
Genetics Department
University of Wisconsin
Madison, Wisconsin 53706

LITERATURE CITED

BAILEY, N. T. J., 1961 *Introduction to the Mathematical Theory of Genetic Linkage*. Clarendon Press, Oxford.

BOTSTEIN, D., R. L. WHITE, M. H. SKOLNICK and R. W. DAVIS, 1980 Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. **32:** 314–331.

CARTER, T. C., and D. C. FALCONER, 1951 Stocks for detecting linkage in the mouse and the theory of their design. J. Genet. **50:** 307–323.

CROW, J. F., 1983 *Genetics Notes*, Ed. 8. Burgess, Minneapolis.

FELSENSTEIN, J., 1979 A mathematically tractable family of genetic mapping functions with different amounts of interference. Genetics **91:** 769–775.

FISHER, R. A., 1948 A quantitative theory of genetic recombination and chiasma formation. Biometrics **4:** 1–9.

FISHER, R. A., 1949a *The Design of Experiments*, Ed. 5. Oliver & Boyd, London.

FISHER, R. A., 1949b Note on the test of significance for differential viability in frequency data from a complete three-point cross. Heredity **3:** 215–219.

GOLDGAR, D. E., P. R. FAIN and W. J. KIMBERLING, 1989 Chiasma-based models of multilocus recombination: increased power for exclusion mapping and gene ordering. Genomics **5:** 283–290.

HALDANE, J. B. S., 1919 The combination of linkage values, and the calculation of distance between the loci of linked factors. J. Genet. **8:** 299–309.

KING, T. R., W. F. DOVE, B. HERRMANN, A. R. MOSER and A. SHEDLOVSKY, 1989 Mapping to molecular resolution in the *T* to *H-2* region of the mouse genome with a nested set of meiotic recombinants. Proc. Natl. Acad. Sci. USA **86:** 222–226.

KOSAMBI, D. D., 1944 The estimation of map distance from recombination values. Ann. Eugen. **12:** 172–175.

LANDER, E. S., and P. GREEN, 1987 Construction of multilocus genetic linkage maps in humans. Proc. Natl. Acad. Sci. USA **84:** 2263–2367.

LATHROP, G. M., and J.-M. LALOUEL, 1988 Efficient computations in multilocus linkage analysis. Am. J. Hum. Genet. **42:** 498–505.

LATHROP, G. M., J.-M. LALOUEL, C. JULIER and J. OTT, 1985 Multilocus linkage analysis in humans: detection of linkage and estimation of recombination. Am. J. Hum. Genet. **37:** 482–498.

LYON, M. F., 1990 L. C. Dunn and mouse genetic mapping. Genetics **125:** 231–236.

OWEN, A. R. G., 1950 The theory of genetical recombination. Adv. Genet. **3:** 117–157.

PASCOE, L., and N. E. MORTON, 1987 The use of map functions in multipoint mapping. Am. J. Hum. Genet. **40:** 174–183.

RAO, D. C., N. E. MORTON, J. LINDSTEN, M. HULTÉN and S. YEE, 1977 A mapping function for man. Hum. Hered. **27:** 99–104.

STAHL, F., 1989 The linkage map of T4. Genetics **123:** 245–248.